Congratulations! You passed!

Grade received 80%

Latest Submission Grade 80%

To pass 80% or higher

Go to next item

1. Which of the following are true? (Check all that apply.)

0 / 1 point

- [x] W_1 is a matrix with rows equal to the parameter vectors of the first layer.

! **This should not be selected**
No. The notation convention is that the superscript number in brackets indicates the number of layers.

- [x] $w_3^{[4]}$ is the row vector of parameters of the fourth layer and third neuron.

! **This should not be selected**
No. The vectors $w_k^{[j]}$ are column vectors.

- [ ] $W^{[1]}$ is a matrix with rows equal to the transpose of the parameter vectors of the first layer.
- [x] $W^{[1]}$ is a matrix with rows equal to the parameter vectors of the first layer.

! **This should not be selected**
No. The parameter vectors are column vectors.

- [x] $w_3^{[4]}$ is the column vector of parameters of the fourth layer and third neuron.

✓ **Correct**
Yes. The vector $w_j^{[i]}$ is the column vector of parameters of the i-th layer and j-th neuron of that layer.

- [ ] $w_3^{[4]}$ is the column vector of parameters of the third layer and fourth neuron.

Expand

Incorrect
You didn't select all the correct answers

2. The sigmoid function is only mentioned as an activation function for historical reasons. The tanh is always preferred without exceptions in all the layers of a Neural Network. True/False?

1 / 1 point

- (●) False
- ( ) True

Expand

Correct
Yes. Although the tanh almost always works better than the sigmoid function when used in hidden layers, thus is always proffered as activation function, the exception is for the output layer in classification problems.

3. Which of these is a correct vectorized implementation of forward propagation for layer l, where $1 \leq l \leq L$?

1 / 1 point

- ( ) $Z^{[l]} = W^{[l]}A^{[l]} + b^{[l]}$
  $A^{[l+1]} = g^{[l]}(Z^{[l]})$
- ( ) $Z^{[l]} = W^{[l-1]}A^{[l]} + b^{[l-1]}$
  $A^{[l]} = g^{[l]}(Z^{[l]})$
- (●) $Z^{[l]} = W^{[l]}A^{[l-1]} + b^{[l]}$

  $A^{[l]} = g^{[l]}(Z^{[l]})$

Expand

Correct

4. You are building a binary classifier for recognizing cucumbers (y=1) vs. watermelons (y=0). Which one of these activation functions would you recommend using for the output layer?

1 / 1 point

- ReLU
- sigmoid
- Leaky ReLU
- tanh

Expand

Correct
Yes. Sigmoid outputs a value between 0 and 1 which makes it a very good choice for binary classification. You can classify as 0 if the output is less than 0.5 and classify as 1 if the output is more than 0.5. It can be done with tanh as well but it is less convenient as the output is between -1 and 1.

5. Consider the following code:

1 / 1 point

#+begin_src python

x = np.random.rand(4, 5)

y = np.sum(x, axis=1)

#+end_src

What will be y.shape?

- (4, 1)
- (1, 5)
- (4,)
- (5,)

Expand

Correct
Yes. By using axis=1 the sum is computed over each row of the array, thus the resulting array is a column vector with 4 entries. Since the option keepdims was not used the array doesn't keep the second dimension.

6. Suppose you have built a neural network. You decide to initialize the weights and biases to be zero. Which of the following statements is true?

1 / 1 point

- Each neuron in the first hidden layer will perform the same computation in the first iteration. But after one iteration of gradient descent they will learn to compute different things because we have "broken symmetry".
- Each neuron in the first hidden layer will compute the same thing, but neurons in different layers will compute different things, thus we have accomplished "symmetry breaking" as described in the lecture.
- The first hidden layer's neurons will perform different computations from each other even in the first iteration; their parameters will thus keep evolving in their own way.
- Each neuron in the first hidden layer will perform the same computation. So even after multiple iterations of gradient descent, each neuron in the layer will be computing the same thing as other neurons.

Expand

Correct

7. Using linear activation functions in the hidden layers of a multilayer neural network is equivalent to using a single layer. True/False?

1 / 1 point

- False
- True

Expand

Correct
Yes. When the identity or linear activation function $g(c) = c$ is used the output of composition of layers is equivalent to the computations made by a single layer.

8. Which of the following is true about the ReLU activation functions?

1 / 1 point

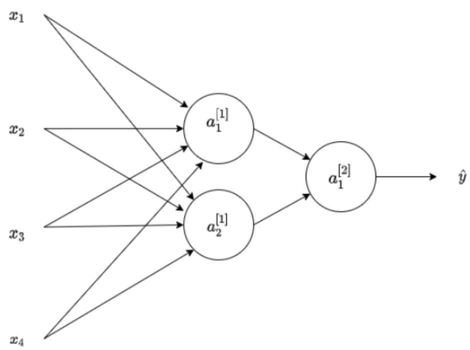
- (selected) They are the go to option when you don't know what activation function to choose for hidden layers.
- They are increasingly being replaced by the tanh in most cases.
- They cause several problems in practice because they have no derivative at 0. That is why Leaky ReLU was invented.
- They are only used in the case of regression problems, such as predicting house prices.

Expand

Correct

9. Consider the following 1 hidden layer neural network:

1 / 1 point



Which of the following statements are True? (Check all that apply).

- [ ] $W^{[2]}$ will have shape (2, 1)
- [x] $W^{[1]}$ will have shape (2, 4).

  Correct

  Yes. The number of rows in $W^{[k]}$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.

- [ ] $W^{[1]}$

  will have shape (4, 2).
- [x] $$b^{[1]}$$ will have shape (2, 1).

  Correct

  Yes. $$b^{[k]}$$ is a column vector and has the same number of rows as neurons in the k-th layer.

- [x] $$W^{[2]}$$ will have shape (1, 2)

  Correct

  Yes. The number of rows in $$W^{[k]}$$ is the number of neurons in the k-th layer and the number of columns is the number of inputs of the layer.

- [ ] $$b^{[1]}$$ will have shape (4, 2)

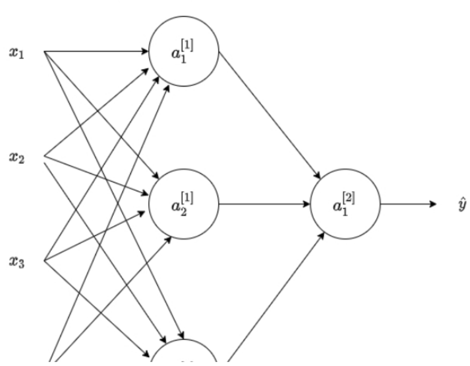
Expand

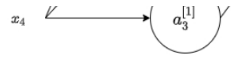
Correct

Great, you got all the right answers.

10. Consider the following 1 hidden layer neural network:

0 / 1 point





What are the dimensions of $Z^{[1]}$ and $A^{[1]}$?

- ◯ $Z^{[1]}$ and $A^{[1]}$ are (4, 1)
- ◯ $Z^{[1]}$ and $A^{[1]}$ are (3, m)
- ◉ $Z^{[1]}$ and $A^{[1]}$ are (3, 1)
- ◯ $Z^{[1]}$ and

Expand

Incorrect

No. The $Z^{[1]}$ and $A^{[1]}$ are calculated over a batch of training examples. The number of columns in $Z^{[1]}$ and $A^{[1]}$ is equal to the number of examples in the batch, m. And the number of rows in $Z^{[1]}$ and $A^{[1]}$ is equal to the number of neurons in the first layer.